YOLO - ASSIGNMENT 9

Mirko Morello
920601
m.morello11@campus.unimib.it

Andrea Borghesi
916202
a.borghesi1@campus.unimib.it

May 28, 2024

1 INTRODUCTION

In this assignment, we explore and analyze the performance of the You Only Look Once (YOLO) [3] object detection architecture on some free-licensed videos. The YOLO model was trained on the Vehicles-OpenImages dataset to detect five classes of vehicles: Ambulance, Bus, Car, Motorcycle, and Truck. We evaluate the trained model's ability to correctly identify and localize these objects in the video frames and analyze any detection errors.

2 TECHNOLOGIES USED

2.1 You Only Look Once (YOLO)

YOLO is a state-of-the-art, real-time object detection system. It reformulates object detection as a single regression problem, directly predicting bounding boxes and class probabilities from full images in one evaluation. This unified architecture enables end-to-end training and real-time speeds while maintaining high average precision.

2.2 Architecture Details

The YOLO model used in this assignment is YOLOv5s, a small variant of the YOLOv5 architecture. It consists of a backbone network for feature extraction, followed by a neck for feature aggregation and a head for bounding box regression and classification. The model was trained for 50 epochs on the Vehicles-OpenImages dataset.

3 DATA

3.1 Training Data

The model was trained on the Vehicles-OpenImages dataset, which contains images of vehicles annotated with bounding boxes and class labels. The dataset includes five classes: Ambulance, Bus, Car, Motorcycle, and Truck. The data was split into training and validation sets.

3.2 Evaluation Data

For evaluation, we fetched some free-licensed videos [1][2] containing instances of one or more of the five vehicle classes. The video was processed frame by frame to assess the model's performance on real-world data.

4 RESULTS

Figure 1: Example of a false positive detection, classifying a two close cars as a truck

Figure 2: Example of false negative detection, the image is too cluttered and no cars are detected

4.1 Qualitative Analysis

The YOLOv5s model demonstrates varying performance across different contexts and settings, as evidenced by the three images. In Figure 2 and Figure 1, which depict busy street scenes with numerous vehicles, the model successfully detects and localizes only a few prominent vehicles, such as trucks and buses. This showcases the model's ability to identify larger vehicles in complex, real-world scenarios. However, in these cluttered scenes, the model struggles to detect smaller vehicles, particularly cars in the background or those partially occluded by other objects. This highlights a common limitation of object detection algorithms in handling small objects in dense environments. In contrast, Figure 3 presents a different context, likely captured from a vehicle's perspective rather than from above. In this less cluttered scene with fewer objects and a more representative angle, the YOLOv5s model accurately identifies all the vehicles present, including cars and an ambulance. This suggests that the model performs better when the input data closely resembles the training data in terms of object scale, viewing angle, and scene complexity.

Figure 3: Example of a proper detection

4.2 Error Analysis

The YOLOv5s model exhibits several types of errors across the three images:

- **False Positives**: In Figure 1, the model incorrectly classifies two cars as a single truck, possibly due to the visual similarity between the side profiles of cars and trucks when they are partially occluded or close together.
- **Missed Detections**: In Figure 1 and Figure 2, the model fails to detect several cars, particularly those in the background or partially occluded by other vehicles. This could be attributed to the small size of these objects and the model's difficulty in distinguishing them from the background clutter in dense scenes.
- **Incorrect Bounding Boxes**: Some of the predicted bounding boxes in Figure 1 are not tightly aligned with the actual vehicle boundaries, indicating challenges in accurately regressing the bounding box coordinates, especially for objects with complex shapes or occlusions.

However, in Figure 3, the model demonstrates accurate detection and localization of all vehicles present, with no noticeable errors. This suggests that the model's performance is highly dependent on the characteristics of the input data, such as object scale, viewing angle, and scene complexity. To address these limitations and improve the model's performance, several approaches can be explored, such as data augmentation to include more diverse and challenging examples, using multi-scale feature representations, incorporating context information, or adapting the model architecture to better handle small objects and complex scenes. Overall, the YOLOv5s model demonstrates promising performance in vehicle detection tasks, particularly for larger vehicles and in scenarios that closely resemble the training data. However, its limitations in detecting small objects in cluttered scenes and its sensitivity to input data characteristics highlight areas for further improvement, that can be achieved by using a more modern versions than the 5s or a more complex size of YOLO, as the one we used was the smallest one available.

References

[1] Ambulance. Link, 2023. Accessed: 2024-05-28.

[2] Cars, busy streets, city traffic. Link, 2023. Accessed: 2024-05-28.

[3] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection, 2016.